

# ON THE TYPOLOGY OF ZERO ANAPHORA\*

C.T.James Huang

## 1. Introduction

Languages differ considerably in the extent to which they allow the use of zero pronouns. The occurrence of zero pronouns ranges from very limited (as in English and French), to somewhat less so (as in Italian, Spanish, etc.), to very free (as in Chinese, Korean, Japanese, etc.). In the first kind of languages, a zero pronoun may occur only as the subject of a tenseless clause, but not as the subject of a tensed clause or as the object of any clause. This is illustrated by the English examples below, where *e* marks the position of a zero pronoun:

- (1) a. John promised Bill [*e* to see Mary].  
b. John preferred [*e* seeing Mary].
- (2) a. \*John promised Bill that [*e* would see Mary].  
b. \*John promised Bill that [Mary would see *e*].  
c. \*John promised Bill [to see *e*].  
d. \*John preferred [Mary's seeing *e*].

This restriction appears to be purely grammatical in nature, having nothing to do with semantic or pragmatic factors. This is clear from the following discourse. Although the reference of an otherwise omitted pronoun is clear, omission is prohibited:

- (3) Speaker A: Did John see Bill yesterday?

- Speaker B:
- a. Yes, he saw him.
  - b. \*Yes, *e* saw him.
  - c. \*Yes, he saw *e*.
  - d. \*Yes, *e* saw *e*.
  - e*. \*Yes, I guess *e* saw *e*.
  - f. \*Yes, John said *e* saw *e*.

---

\*An earlier version of this paper was presented at the 17th Language Research Conference of the Language Research Institute, Seoul National University, in December 1983. I am grateful to the participants who offered their generous comments and suggestions. In addition, of their kind assistance and encouragements in various ways I am particularly indebted to Professors Han-kon Kim, Chungmin Lee, Dong-Whee Yang, and In-Seok Yang. Much of the material used in this paper is taken from a longer paper of mine (Huang 1984), where some of the issues raised here are given a more detailed treatment.



- know that saw  
 'José knows that [he] saw Maria.'  
 b. José sabe [que Maria *e viu*].  
     know that saw  
 'José knows that Maria saw [him].'

The wide variety among languages concerning the distribution of a zero pronoun has drawn the attention of many researchers in recent years. The most important problem that arises for the theory of Universal Grammar (UG) is to formulate an appropriate parameter (or parameters) for the optimal characterization of how languages should differ in precisely the way they do. A number of important works in generative grammar have been devoted to this problem (e.g. Perlmutter 1971, Borer 1983, Jaeggli 1982, Chomsky 1981, Taraldsen 1978, McClosky and Hale 1983, Rizzi 1982 among many others). This paper represents yet another attempt at finding a solution to this problem.

This paper is arranged as follows: In section 2, I will review an important and by now well known theory, represented by work of Chomsky and others, and consider its explanatory power as well as its scope and limitations. In section 3, I will propose an alternative in summary form. The rest of the paper will substantiate the claims made in section 3 and indicate the empirical and theoretical consequences of the proposal.

## 2. The Agreement-Based Theory

The difference between the English-type languages on the one hand and the Italian-type and Chinese-type Languages on the other lies in whether a language allows a zero pronoun in the subject position of a tensed clause. This difference has been assumed to be the result of what has come to be known, descriptively, as the "Pro Drop Parameter" (cf. Chomsky 1981) or the "Null Subject Parameter" (cf. Rizzi 1982). One important type of explanation that has been proposed to derive this parameter, in particular as it applies to distinguish between English-type and Italian-type languages, is based on the idea of recoverability and the observation, due to Taraldsen (1978), that the possibility of pro drop in a language often correlates with the existence in it of a rich system of agreement. According to this theory, as assumed in Chomsky (1981, 1982), Italian and Spanish allow a subject pronoun to drop from a tensed clause, because they each have a rich verb-subject agreement system. The agreement marking on a verb is rich enough to determine (and in some sense, recover) the reference of a missing subject; therefore such a subject may drop. On the other hand, in English and French the agreement system is somewhat degenerate, and the agreement marking on a verb is too meager to determine the reference of a missing subject; therefore pro drop may not occur. Furthermore, since neither types of languages exhibit any verb-object agreement, no object pronoun may drop in any of these languages.

This explanation appears to be quite plausible, and is further supported by an

important piece of evidence from Pashto, a split ergative language.<sup>1</sup> Pashto uses an accusative agreement system for sentences in the present tense: the verb agrees with the subject whether transitive or intransitive. But for sentences in the past tense, agreement goes ergative: the verb agrees with the subject if intransitive, but with the object if transitive. The following sentences illustrate the present-past contrast in agreement in Pashto transitive sentences:

- (7) a. zə   mana   xwr-əm.  
           I    apple eat-1msg  
           'I eat the apple.'  
       b. ma   mana   wə-xwar-a.  
           I    apple PRF-eat-3fsg  
           'I ate the apple.'

Precisely as the agreement-based theory would predict, only the subject may drop in (7a), but only the object may drop in (7b):

- (8) a. e    mana   xwr-əm.  
           apple eat-1msg  
           '[I] eat the apple.'  
       b. ma e   wə-xwar-a.  
           I    PRF-eat-3fsg  
           'I ate [it (fem.)].'

If the object in (8a) or the subject in (8b) were omitted instead, the sentences would be ungrammatical:

- (9) a. \*zə e xwr-əm.  
           I    eat-1msg.  
           'I eat [?].'  
       b. \*e   mana w-xwar-a.  
           apple PRF-eat-3fsg  
           '[?] ate the apple.'

The possibility of pro drop, as well as the distribution of the zero pronoun, is clearly dictated by the presence of agreement. Pashto thus provides very important support for the theory based upon the Taraldsen generalization.

However, this theory runs into difficulty in a number of areas. First, it does not account for the fact that in the vast number of languages of the third type comprising Chinese, Japanese, Korean, etc., a subject pronoun may drop freely in tensed clauses. It is well known that these languages do not exhibit any agreement whatsoever. Secondly, it does not explain why a zero subject pronoun is possible in the non-finite clauses of all languages. Since non-finite clauses generally do not contain agreement, it is not clear why the occurrence of a zero subject in them is not

---

<sup>1</sup>I am indebted to Farooq Babrakzai for help with the Pashto data.

subject to the same requirement.<sup>2</sup> Thirdly, in Chinese-type languages even objects (and other non-subjects) can drop, again in the absence of any agreement marking. The agreement-based theory again does not extend to this fact. (For Portuguese, although the occurrence of a null subject in a finite clause may be attributed to the presence of agreement, the occurrence of a null object as in (6b) also presents the same problem.)

We thus have a theory that must be right in some cases but must be wrong in others. Problems of the kind we have just indicated have led certain researchers in the field to doubt the relevance of agreement, and to abandon the agreement-based theory altogether. For example, Gundel (1980), based upon considerations similar to those mentioned here, claims that the typology of zero anaphora has nothing to do with agreement, but that the observed differences among languages should be derived as a consequence of the “topic-prominence” typology originally proposed by Li and Thompson (1976). Gundel proposes, in very general terms, that the more “topic prominent” a language is, the more likely it is to drop a pronoun; and the more “subject prominent” it is, the less likely it is to drop a pronoun. Chinese, Japanese, Korean, etc. are said to be “topic prominent”, so the fact that they allow a maximal use of zero pronouns follows from the general principle. English, on the other hand, is “subject prominent”, so it exhibits strict restrictions on the occurrence of zero pronouns. There appears to be some plausibility in Gundel’s approach, but it should be easy to see that the theory leaves many more problems than it is purported to solve. For the sake of brevity, let us mention only two of the most obvious problems. First, the theory completely ignores the relevance of agreement as evidenced by data of the kind provided by Pashto. Secondly, the theory does not make any predictions when questions of any detail are considered. For example, the theory does not tell us why the notion of topic prominence will derive the exact patterns we have observed, or why English allows a zero pronoun in the subject position of a tenseless clause but not in a tensed clause, and why the situation is not the other way around. Furthermore, it is not clear why the “subject prominence” of English has anything to do with its not allowing a zero object pronoun. As we shall see later, moreover, there are important asymmetries between subject and object even in languages of the Chinese type which the theory does not seem capable of explaining. It is apparent that a more nearly adequate theory is needed.

### 3. Outline of Proposed Analysis

I would like to argue that a proper solution to the problems raised here requires that we look at the facts so far considered as instances of two distinct facts, namely the occurrence of a zero subject and the occurrence of a zero object, and that differences among languages with respect to these two facts should be dealt with separately, each involving an independent typological parameter.

---

<sup>2</sup>In Chomsky (1981) a PRO appearing as the subject of an infinitive is assumed to be a pronominal anaphor whose reference is governed by the control theory. The question is why the agreement-based theory must be stipulated to affect only pronominal *non*-anaphors.

The occurrence of a zero *subject* pronoun will be considered the proper subject matter for the theory of pro drop. Because a zero subject pronoun can occur where there is a rich agreement marking (as in Italian, Spanish, Pashto, etc.) and where there is no agreement marking whatsoever (as in finite clauses of Chinese-type languages and in non-finite clauses of all languages), a paradox arises for the agreement-based theory. I propose that this paradox can be resolved if the theory is generalized to allow for two possible ways in which the reference of a zero pronoun can be determined, in accordance with the condition of recoverability: either by a rich enough AGR, or by an antecedent NP. This general theory, in effect, subsumes the agreement-based theory and Chomsky's (1980) theory of control, treating both AGR and full NPs as both possible "controllers" of empty pronouns. In section 6 below, I will give more substance to this proposal, and show how the generalized theory of control can properly derive the facts associated with the pro drop parameter.

As for the occurrence of a zero object, I will argue that such an empty category (EC) is not a pronominal, but rather a variable in the sense of Chomsky (1981: 330):<sup>3</sup>

- (10) a. An EC is pronominal if and only if it is free or locally bound by an element with an independent thematic role, and a non-pronominal otherwise.
- b. A non-pronominal EC is an anaphor if and only if it is locally A-bound, and a variable if locally A'-bound.

More specifically, we claim that an object EC of the kind seen in Chinese, etc., is locally bound by an empty topic or an abstract operator of some sort. Thus, for a sentence like (11a) in Chinese, the proper representation is (11b), where OP marks the position of an empty operator:

- (11) a. Zhangsan shuo Lisi kanjian *e* le.  
           say      see      ASP  
           'Zhangsan said that Lisi saw [him].'
- b. [OP, [Zhangsan shuo Lisi kanjian *e*, le]].

The object EC is treated on a par with a trace of wh-movement, and the structure (11b) is on a par with that of a topicalized sentence like (12):

- (12) neige xiaohai, [Zhangsan shuo Lisi kanjian *e*, le].  
       that child          say      see      ASP  
       'That child, Zhangsan said that Lisi saw.'

Because the object EC is a variable and not a pronominal according to the definition in (10), it falls outside of the domain of the theory of pro drop. The question remains, of course, as to what distinguishes the Chinese type of languages from

<sup>3</sup>A node A is bound iff it is c-commanded by a coindexed antecedent. Otherwise it is free. A node is A-bound iff the binder occurs in argument position (subject, object, etc.), and A'-bound if the binder occurs in operator position (COMP, topic, head, etc.).

the other two types with respect to the possible occurrence of an object EC bound by an *empty* operator. I suggest that this difference can be made to follow from a separate parameter, which can be shown to have other consequences. Section 5 will discuss this parameter in more detail.

Another question this conception of the object EC raises is why an object EC cannot be a pronominal. As I will show in section 6, an answer to this question is available from a proper formulation of the pro drop parameter and other independent principles of grammar, in particular the binding theory of Chomsky (1981, 1982).

The theory proposed will thus recognize two independent parameters for a proper account of the distribution of what appears to be a zero pronoun in language. This typological scheme yields four possible language types, depending on whether a given language is a pro drop language or not, and whether it is an “empty topic” language or not. We shall see shortly, in section 4, that each of the four types is exemplified by actual language data. The first thing I will do, however, is to provide arguments for the claim that an object EC is always a variable.<sup>4</sup>

#### 4. The Status of Empty Objects

I will offer three pieces of evidence for the claim that an object EC of the kind occurring in Chinese-type languages is a variable but not a pronominal. The first two will show that such an EC exhibits properties of a variable, and not those of a pronominal, consistent with general principles of grammar. Two well known properties of a variable relevant for our discussion are (a) that it may not be A-bound, and (b) that it may be A'-bound (cf. Chomsky 1981, May 1977). An object EC can be shown to have both these two properties. Consider (13) and (14):

- (13) Zhangsan shuo [*e* bu renshi Lisi].  
           say           not know  
       ‘Zhangsan said that [he] doesn’t know Lisi.’
- (14) Zhangsan shuo [Lisi bu renshi *e*].  
           say           not know  
       ‘Zhangsan said that Lisi doesn’t know [him].’

Each of the ECs in these sentences is translated into English with an overt pronoun. However, there is an important difference in interpretation between the two sentences that the English translation does not reflect. In particular, although the embedded subject EC in (13) can be interpreted as taking the matrix subject as its antecedent or as referring to someone else whose reference is fixed in discourse (the “discourse topic”), the embedded object EC in (14) can only be interpreted as referring to a discourse topic, but not as being referentially bound by the matrix subject

---

<sup>4</sup>I will not consider NP-traces (which may occur in object position but are anaphors). Also I will exclude from further consideration languages like Pashto. An object EC in a past tense transitive sentence in Pashto can, of course, be a pronominal, like a subject EC in Italian or any other pro drop languages.

*Zhangsan*. This restriction does not apply, of course, to the pronoun *him* in the English translation. This is also not a property peculiar to an overt object pronoun in Chinese, as the following sentence is clearly ambiguous:

- (15) Zhangsan shuo [Lisi bu renshi ta].  
           say          not know he  
       ‘Zhangsan said that Lisi doesn’t know him.’

The asymmetry between (13) and (14) can be observed in other languages of the third type exhibiting object ECs. The following sentences from Portuguese (Wynn Chao, personal communication) exhibit the same contrast in the interpretation of the EC:

- (16) João sabe que e gosteria de conhecer a Maria melhor.  
           know that would-like know better  
       ‘João knows that [he] would like to know Maira better.’
- (17) João sabe que a Maria gosteria de conhecer e melhor.  
           know that would-like know better  
       ‘João knows that Maira would like to know [him] better.’

Only the subject EC in (16) but not the object EC in (17) can be interpreted as being bound by the matrix subject *João*.

We have seen that an object EC cannot be A-bound by a matrix subject. (In fact, such an EC cannot be bound to any other A-position either, though no examples have been given to show this.) If an object EC were a pronominal, then there would be no reason to expect that it cannot be A-bound. This property is expected, however, if we assume that such an EC is a variable. As a variable, such an EC may be expected to be A'-bound by an operator. This expectation is fulfilled if an A'-binder is present:

- (18) neige ren<sub>i</sub>, Zhangsan shuo [Lisi bu renshi e<sub>i</sub>].  
           that man          say not know  
       ‘That man, Zhangsan said that Lisi doesn’t know.’

A word of clarification is in order here before we go further. In claiming that an object EC may not be A-bound, I mean that such an EC may not be interpreted as referentially dependent upon a c-commanding NP in argument position for its reference. This restriction does not even imply that such an EC cannot be *coreferential* with a c-commanding argument. A pronominal may come to be coreferent with another element by picking up its reference either (a) from the latter element, thus taking it as its linguistic antecedent, or (b) from elsewhere. The second situation happens when the reference of the anaphoric element is inferred from discourse context. In this situation the reference is often determined on the basis of non-linguistic factors. Consider an example of such a situation:



(19) Speaker A: *shei bu renshi Zhangsan?*

who not know

'Who doesn't know Zhangsan?'

Speaker B: *Zhangsan shuo [Lisi bu renshi e].*

say not know

'Zhangsan said Lisi doesn't know [him].'

The sentence in speaker B's reply is exactly the same in form as (14). The object EC here can refer to the matrix subject, unlike that in (14). This interpretation is possible, of course, only because of the context. There is no reason to assume that the EC has to pick up its reference from the matrix subject directly. Its reference can be inferred from context. Assuming that speaker B obeys the conversational principle of cooperation, it is most natural to assume that the EC refers to the person that speaker A asks about, namely Zhangsan. In this case, the object EC may be said to have no linguistic antecedent, or, if it has an antecedent, that its antecedent is the *Zhangsan* mentioned in speaker A's question, not its matrix subject. It is important to note that our claim about the contrast between (13) and (14) is made only with respect to the possibility of a given EC taking the matrix subject as a linguistic antecedent. It is assumed therefore that they are uttered completely out of context, so that there is no other way to infer the reference of the EC. The significance of this contrast becomes clearer if we replace the ECs in (13) and (14) each with an overt pronoun, where no contrast whatever is observed.

Not only is it important to exclude contextual factors in order to see the point we are making, it is also necessary that the sentences used represent pragmatically neutral situations so that the role of pragmatic inference is reduced to a minimum. Consider the following sentence:<sup>5</sup>

(20) *xiaotou yiwei [mei ren kanjian e], suoyi ....*

thief think no man see so

'The thief thought that no one saw [him], so .... (e.g. took the package and ran).'

The object EC can be construed as referring to 'the thief'. There is no reason to assume, however, that it depends upon the matrix subject as its linguistic antecedent. The meaning of the sentence is such that the reference of the EC can be directly inferred to be the thief. In such a sentence, even an overt pronoun is almost always interpreted as referring to the thief:

(21) *xiaotou yiwei [mei ren kanjian ta], suoyi ....*

thief think no man see he so

'The thief thought that no one saw him, so ....'

The possibility of the overt pronoun referring to someone other than the matrix sub-

<sup>5</sup>Xu (1984) uses sentences like (20) to argue that an object EC may sometimes be A-bound. It seems this misses the point. The fact remains that in such sentences as (14) the EC cannot be A-bound, in the sense that an overt pronoun in its place can, and in the sense that the subject EC in (13) also can.

ject in (21) is much lower than in the pragmatically more neutral sentence (15).

We have seen that an object EC behaves more like a variable than a pronominal in that it may be A'-bound by a topic but not A-bound by a subject. Another piece of evidence that shows the same point is the fact that it may be A'-bound by the head of a relative clause (i.e. relativized), but not A-bound by something else. Consider the following sentence, where the relative clause contains a subject EC and an object EC:

- (22) Mali hai zhao-bu-dao yige [*e* xihuan *e* de] nanren.  
 Mary still can't-find one like RM man  
 'Mary still can't find a man who [she] likes.'

Consider only the situation where one of the ECs is bound by the matrix subject 'Mary', and the other EC relativized by 'man'. There are two logical possibilities: either (a) the subject EC is A-bound (as a pronominal) by 'Mary' and the object EC relativized (as a variable) by 'man', or (b) vice versa. In the situation (a), the sentence would mean "Mary still can't find a man who she loves", and in the situation (b), it would mean "Mary still can't find a man who loves her". However, the sentence is not ambiguous: it only has the first interpretation, where the subject EC is a pronominal and the object EC is a variable. There is no reason why a subject cannot be relativized as a variable in general. But if we assume that an object EC cannot be a pronominal, then the non-ambiguity of (22) is automatically accounted for.

A similar piece of evidence is available from free relatives. The following sentence contains a coordinate subject with two free relatives each of which contains two ECs:

- (23) [*e* mai *e* de] gen [*e* zu *e* de] dou hao.  
 buy RM and rent RM all good  
 'What one buys and what one rents are both good.'

For each free relative, consider the case where one of the ECs is relativized and the other used as an arbitrary (generic) pronoun referring to people or things in general (like arbitrary PRO). Each free relative represents two logical possibilities. The first free relative could mean either "what one buys" or "the person that buys things (= a buyer)". The second free relative could also mean either "what one rents" or "the person that rents things (= a renter)". The sentence with the two free relatives could then be four-way ambiguous. However, as indicated in the translation, it has only one reading. This is precisely the situation where the object ECs are each relativized as a variable. The assumption that an object EC cannot be a pronominal correctly excludes all the other three readings.

As proposed in Chomsky (1981), the fact that a variable cannot be A-bound follows from the binding theory and the assumption that a variable is an R-expression, on a par with other R-expressions including names like *John*, *the man*, and anaphoric epithets like *the sissy*, *the bastard*. It is well known (since Lasnik 1976) that names and anaphoric epithets cannot be A-bound:

- (24) a. \*John loves John's mother.

- b. \*John said that John went to the party.  
 (25) a. \*John loves the sissy's mother.  
 b. \*John said that the sissy went to the party.

It seems that names and anaphoric epithets also share with variables the property that they may be A'-bound:

- (26) a. John, I like John.  
 b. John, I like the sissy.

Another well known property of names and epithets is that they may be coindexed with NPs in argument position as long as these NPs do not c-command them (Lasnik 1976):

- (27) a. John's mother loves John.  
 b. When I saw John, John was leaving.  
 (28) a. When I saw John, the sissy was leaving.  
 b. When I saw the sissy, John was leaving.

If we assume that an object EC is a variable, one may expect that it also shares this property with the other R-expressions. The following sentences show that this is indeed the case in Chinese:

- (29) wo mei kan *e*, ta jiu ba shu na zou le.  
 I not read he then BA book take away ASP  
 'Before I read [it], he took the book away.'  
 (30) renshi *e* de ren dou shuo Zhangsan shi yige hao ren.  
 know RM man all say is one good man  
 'Everyone that knows [him] says Zhangsan is a good man.'

In short, the referential properties of an object EC we have seen show that it is more like an R-expression than a pronominal. An empty R-expression, in the framework of Chomsky (1981), is a variable, and the properties of the object EC are largely consistent with the assumption that it is indeed a variable. There are two obvious differences between the object EC in Chinese and a variable as commonly observed in English. The first is shown by the fact that the English counterparts of sentences like (29) and (30) are ill-formed:

- (31) \*Before I read *e*, he took the book away.  
 (32) \*Everyone that knows *e* says that John is a good man.

The other difference is that sentences having the form of (14) are also ill-formed in English even with the single reading that the object EC refers to a discourse topic:

- (33) \*John said Bill doesn't know *e*.

One obvious fact about these sentences in English and Chinese is that they involve an object EC that is apparently not bound on the surface. A plausible way to ac-

count for the typological difference between English and Chinese is to directly stipulate that in Chinese-type languages a variable need not be A'-bound. An indication of the plausibility of this hypothesis is the fact that in such a language often an operator need not literally bind a variable either. (34) is an example of this well known fact:

- (34) neichang huo, xingkui      xiaofangdui      lai      de      zao.  
       that    fire    fortunately fire-brigade    come COMP    early  
       'As for that fire, fortunately the fire brigade came early.'

One may thus assume that in these languages the general principles of quantification need not obtain: an operator may be vacuous, and a variable may be free. However, if nothing further is said, this idea would lead us to expect that a variable could in principle refer to just anything, as far as the antecedent does not A-bind it. The fact is that even in (14), the variable is required to refer to a discourse topic, not just anything imaginable. Furthermore, when there is an overt topic (as in (15)), the variable must be interpreted as bound by that topic. This suggests a different way of looking at the difference between English-type and Chinese-type languages: instead of saying that the variable in (14) is unbound, I will assume that it is bound by an empty topic. Thus, (14) would be represented on a par with (15), except that its topic is not lexically represented as in the latter. An important piece of evidence for this assumption comes from German, as the following fact shows:<sup>6</sup>

In spoken German a subject or an object pronoun may be deleted optionally. Thus, given a sentence like (35), one may omit the subject, (as in (36)), or the object (as in (37)):

- (35) Ich habe ihn    schon    gekannt.  
       I    have him    already    known  
       'I already knew him.'  
 (36) Habe ihn schon gekannt.  
 (37) Habe ich schon gekannt.

An important restriction on this phenomenon of pronoun drop is that a pronoun may be omitted only from the topic position, i.e. the first position within a sentence. This is shown by the fact that in each of (36) and (37) the verb must occur sentence-initially on the surface. Given the well known requirement that a main verb occurs in second position in this language, it is obvious that the missing pronoun must have occurred in first position before it got deleted. If a pronoun gets deleted directly from an argument position, the result is ungrammatical, as the following shows:

- (38) \*Ich habe *e* schon gekannt.  
 (39) \*Ihn habe *e* schon gekannt.

Since the first and the second positions are already occupied, the missing NP can-

<sup>6</sup>The German facts reported here are due to Ross (1982) and Irene Heim (personal communication).

not have been deleted from the topic position, and the sentences are ill-formed. These sentences show that an object (as well as a subject) must be moved to topic position before it gets deleted, within a framework that assumes movement for topicalization. What looks like an object pronoun is thus in fact a variable bound by a topic. This is exactly how I have proposed to look at object ECs in Chinese-type languages. German thus not only provides another piece of evidence for the claim that an object EC is a variable, but also lends clear support for the hypothesis that such a variable is not unbound, but rather bound by an empty operator.

Another fact about German pronoun drop is that only one pronoun may drop per sentence. Thus, if both the subject and object are deleted from (35), the result is ill-formed:

(40) \*habe schon gekannt.

This fact may be considered a consequence of the fact already observed, that every pronoun must be moved into topic position before it gets deleted, plus the fact that there is only one topic position per sentence in this language, the verb being always in second position.

I have argued in this section that an object EC cannot be a pronominal. Rather such an EC is a variable bound by an operator, which may be empty in languages of the Chinese type, but not in languages of the English or of the Italian type. I will assume that this typological difference arises from a parameter independent of the pro drop parameter. This typological scheme yields four possible language types. English-type languages are neither pro-drop nor empty-topic languages. Italian-type languages, on the other hand, are pro-drop, but nonempty-topic, languages. Chinese-type languages are both pro-drop and empty-topic, since an object EC may exist as a variable bound by an empty topic (as in (14)), and a subject EC may exist as a pronominal bound by a matrix subject (as in (13)). The fourth type of language, empty-topic but non-pro-drop, is represented by German. The grammaticality of sentences with a pronoun deleted from the first position shows that it is an empty-topic language. Furthermore, the fact that only one pronoun per sentence may be missing shows that there is only way to drop a pronoun, the "empty topic" way, and therefore that it is not a pro-drop language. All four possible types of languages are exemplified. This fact argues, of course, for the recognition of two independent parameters as I have proposed.

## 5. The Empty Topic Parameter

It is doubtful that what we have called the empty topic parameter represents an isolated fact in language. If this is indeed a genuine typological parameter, it is natural to expect that there are correlations to the difference between having and not having an empty topic binding a variable. In this section I will enumerate some of the facts that appear to be characteristic of empty topic languages.

The first fact is the existence of what Tsao (1977) calls "identical topic deletion", which operates across discourse to delete the topic of a sentence under identity with

a topic in a preceding sentence. The result of such a process is formally a "topic chain". An example of a topic chain is given below:

- (41) [Zhongguo, difang hen da.] [*e*, renkou hen duo.]  
       China place very big population very many  
       [*e*, tudi hen feiwo.] [*e*, women dou hen xihuan.]  
       land very fertile we all vary like

'China, (its) land area is very large. (Its) population is very big. (Its) land is very fertile. We all like (it).'

Each of the ECs above marks the site of a deleted topic. This is, of course, what we have been calling an empty topic. One might, in fact, assume that the existence of empty topics derives itself from the existence of the rule of identical topic deletion. In an interpretive framework, this rule may be analyzed as a rule of predication that applies at the LF' level (following LF) which coindexes an empty topic with a discoursally identified topic. A language having this discourse interpretation rule thus allows empty topics, but not a language not having this rule, presumably because in the latter case the reference of an empty topic cannot be determined.

Another characteristic of empty topic languages is the phenomenon of "topic-prominence" (Li and Thompson 1976). A manifestation of this phenomenon is the fact that the topic appears to be a basic unit of a sentence, which cannot plausibly be derived from some non-topic constituent of the sentence. The sentence (34) given above illustrates this fact. One plausible way to formally characterize the "topic-prominence" of Chinese and the "subject-prominence" of English is to say that while Chinese has the PS rules (42 a-b), among others, English has the rules (43 a-b):

- (42) a.  $S' \rightarrow \text{Top } S$   
       b.  $S \rightarrow (\text{NP}) \text{ VP}$   
       (43) a.  $S' \rightarrow (\text{Top}) S$   
       b.  $S \rightarrow \text{NP VP}$

That is, topic is an obligatory element of a clause in Chinese but not subject, while the reverse holds in English. The requirement that clauses have subjects is what Chomsky (1982) has assumed to be part of his Extended Projection Principle. This may be assumed to be a non-universal requirement and taken as a parameter of typology. We may consider it to be a fact about "subject-prominent" languages in Li and Thompson's terms. On the other hand, the requirement that every clause must have a topic may be assumed to be a fact of "topic-prominent" languages.<sup>7</sup> The requirement in English that clauses must have subjects has the well known consequence that pleonastic elements must appear as structural subjects where no other principles of grammar require semantically real subjects. On the other hand, if we assume that every sentence must have a topic in Chinese, then in the absence of an overt

---

<sup>7</sup>Pure existential or presentative sentences, which often serve to introduce topics in a discourse, may be considered to be an exception to this requirement.

topic, an empty topic must be assumed to be present. It is therefore possible that empty-topic languages allow empty topics because such elements are independently required elsewhere in their grammars.

If it is correct to say that the obligatory presence of a topic in Chinese gives rise to empty topics and the obligatory presence of a subject in English gives rise to pleonastic subjects, a natural question arises as to why in English the requirement of a structural subject may not be satisfied by the use of an *empty* pleonastic subject:

- (44) \*[*e*] seems that John won't come tomorrow.

Furthermore, since topics are also possible (though optional) elements in English, what is the principle that rules out the presence of an empty topic in such a language? This is the same question as asking why the "identical topic" interpretation rule cannot exist in English at all to allow for empty topics. A more fundamental solution to the problem of properly formulating the empty topic parameter is suggested by O. Jaeggli (personal communication). Suppose that the following statement holds of Chinese-type but not of English- or Italian-type languages:

- (45) The INFL is a proper governor.<sup>8</sup>

If we assume that all empty categories (except PRO in infinitival subject position) are subject to the ECP, which requires them to be properly governed, then we may explain the impossibility of an empty expletive in English (as in (44)), because the subject EC is not properly governed, INFL not being a proper governor in this language, in violation of the ECP. Similarly, the fact that the language does not allow empty topics also follows from the same account, for the topic is apparently not properly governed if the subject is not.

This account also predicts that pro drop languages like Italian and Spanish do not allow empty topics. The fact that they do allow empty pronominal subjects in tensed clauses can be accounted for along the lines of Chomsky's (1981) and Jaeggli's (1980) original account of the pro drop parameter. Assume that these languages have the option of applying affix hopping in syntax. Then when this happens in syntax, an empty category in subject position is ungoverned. The subject EC is thus a PRO, while the tensed clause is on a par with a tenseless clause at the level after affix hopping has applied. Since subject PRO is not subject to the ECP, its occurrence in these languages is not prevented by this principle.

As for languages like Chinese, the assumption that the INFL is a proper governor has the consequence that the subject is properly governed. It is natural to expect in this case that the topic position is also properly governed. This has the consequence that the topic (as well as the subject) may each be an EC satisfying the ECP, a situation that directly gives rise to the existence of empty topics. We might thus assume that the origin of an empty topic language is the formal feature indicated in (45). This in turn gives rise to the existence of the interpretive rule of

---

<sup>8</sup>In Huang (1982) I assumed that this arises from the INFL in Chinese being lexical. Piccalo (1982) argues that in Catalan the INFL in indicatives is a proper governor, though not the INFL in subjunctives.

“identical topic deletion”. The phenomenon of “topic prominence” may also be assumed to arise from here. The existence of sentences like (34) means that the topic must be Case-marked in its own position. The assumption that that position is properly governed also means that it can be directly Case-marked. On the other hand, in English a sentence like (34) must be rendered in such a way that the topic is preceded by a separate Case marker like *as for*, *concerning*, etc.

(46) As for the fire, fortunately they came in time.

This fact in English may be plausibly assumed to arise from the topic position not being governed by INFL in the language.

Note that our assumption that the topic is properly governed by INFL in empty topic languages predicts that the subject is also properly governed. This is, of course, because the subject is even closer to INFL than the topic. This prediction appears to be correct, a fact that lends considerable support to the assumption. In Chinese (and Japanese and Korean), there appears to be no subject-object asymmetry under extraction (in syntax or in LF) as a manifestation of an ECP effect. The clearest evidence comes from the fact that long movement of a *wh*-in-situ can take place from subject or object position without any difficulty. Thus, the question (47) below may be appropriately answered with (48) or (49).

- (47) ni   xiang-zhidao [shei   maile   shenme]?  
       you wonder       who bought what  
       a. ‘Who is the *x* such that you wonder what *x* bought?’  
       b. ‘What is the *x* such that you wonder who bought *x*?’
- (48) wo xiang-zhidao [Lisi   maile   shenme].  
       I   wonder               bought what  
       ‘I wonder what Lisi bought.’
- (49) wo xiang-zhidao [shei   maile   shu].  
       I   wonder       who bought book  
       ‘I wonder who bought the book.’

On the reading according to which (48) is an appropriate answer, (47) has the following LF representation:

- (50) [Shei<sub>i</sub> [ni   xiang-zhidao [shenme<sub>j</sub> [t<sub>i</sub>   maile   t<sub>j</sub> ]]]].  
       who   you wonder       what               bought

This is on a par with the LF representation of the ill-formed “Who do you wonder what bought?” in English, which is excluded by the ECP. Given (48) as an appropriate answer, (50) must be regarded as well-formed. This shows that the embedded subject trace must be properly governed in accordance with the ECP.

Another piece of support for the hypothesis being entertained comes from Portuguese. In Zubizarreta (1983), it is reported that Portuguese shows no subject-object ECP effects under extraction, and that the language does not exhibit free inversion of the subject. In *pro* drop languages like Italian and Spanish, it has been shown



that they also do not exhibit overt subject-object ECP effects (cf. Rizzi 1982, Jaeggli 1982). However, there is enough evidence that this apparent violation of the ECP is due to the existence of free inversion in these languages. What appears to be long extraction of a preverbal subject is in fact long extraction of an inverted subject from postverbal position, where the long-moved trace is properly governed by the verb. This hypothesis receives further support from the fact that with *wh*'s-in-situ, the extraction that takes place in LF does show a subject-object asymmetry in accordance with the ECP. The situation with Portuguese, however, cannot be settled in the same way, given that the language does not allow free inversion of the subject as Zubizarreta has shown. But as we observed earlier, Portuguese also allows empty topics. These three facts—the lack of subject-object ECP effects, the non-existence of free inversion, and the existence of empty topics—receive a natural explanation, of course, if we only assume that the INFL is a proper governor in Portuguese, as it is in Chinese, etc.

In short, there appears to be a number of properties associated with the property of allowing an empty topic in a given language. A separate parameter is motivated in so far as it can, when properly formulated, provide an explanation for the clustering of these properties.

## 6. The Pro Drop Parameter

Let us turn now to the problem of accounting for the distribution of a genuine empty pronominal. I would like to suggest that a solution is available from a properly formulated theory that determines the reference of such a category. In particular, I propose that empty pronominals (whether they are ungoverned PROs or governed PROs) are subject to a generalized version of the control theory along the lines originally suggested in Chomsky (1980) (cf. Rosenbaum 1967). The theory may be stated in the form of a co-indexing rule:

### (51) The Generalized Control Rule (GCR):

Co-index an empty pronominal with the closest nominal element.

Roughly, an empty pronominal takes the closest potential antecedent as its antecedent. A nominal element will be understood here to mean either NP or AGR. This extended notion of an antecedent is thus intended to tie together the agreement-based theory and the theory of control. I will define "closest" in the following manner. Following Chomsky (1980), A is closer to B than C is if A c-commands B but C does not c-command B. Furthermore, for two nodes A and C both c-commanding B, A is closer to B than C is if A but not C occurs within the same clause as B, or if A is separated from B by fewer clause boundaries than C is.

Basically, what the generalized control rule says is that an empty pronominal either has its reference determined by a rich enough AGR, or is controlled by an NP. Coupled with independent principles of grammar, this conception of control provides a basic account for the distribution and reference of empty pronominals. To see how it works, let us consider sentences of the forms represented below in

various language types.

- (52) a. *e* came.  
 b. John saw *e*.  
 c. John said that *e* saw Bill.  
 d. John said that Bill saw *e*.  
 e. John tried *e* to come.

Since I have proposed the recognition of the empty topic parameter, theoretically each of the ECs above could be a variable bound by an empty operator or it could be a genuine empty pronominal. Let us therefore consider both possibilities.

Consider first a language of the English type. Since English is not an empty topic language (presumably because its topic is not properly governed), the possibility that any of the ECs above is an A'-bound variable is ruled out. So we need only see if any of the ECs can be a genuine pronominal. In (52a), the subject EC as a pronominal is subject to the GCR (51), which requires it to be coindexed with the closest nominal element. The closest nominal element in this case is the AGR contained in the verb *came*, so the GCR will coindex the EC with the AGR. However, because the AGR is too meager in English, it cannot "sufficiently" determine the reference of the EC. The EC thus cannot survive due to the condition of recoverability. Consider now (52b). The GCR requires the object EC to be coindexed with either the AGR or the subject *John*. The AGR is too meager, leaving the subject as the only potential antecedent. However, the EC as a pronominal is also subject to the independent principle that a pronoun must be free in its governing category:

- (53) \**John<sub>i</sub> saw him<sub>i</sub>.*

In the object position, then, an empty pronominal cannot exist because it cannot satisfy either the generalized control theory or the binding theory without violating the other. The situation with (52c) and (52d) is similar. The subject EC in (52c) is required by the GCR to be "identified" by its own AGR which is too meager to do the required job. The object EC in (52d) as a pronominal entails a contradiction of the GCR and the binding theory. Consider now the last case, (52e). The embedded clause does not contain AGR. Therefore, the nominal element closest to the subject EC is the matrix subject. In accordance with the GCR, therefore, it is coindexed with *John*. Since this mode of coindexation does not violate any independent principle of grammar, the subject EC is admitted. In short, in English-type languages, only sentences of the form (52e) are well-formed.

Consider next languages of the Italian type. Because these languages are also by assumption not empty topic languages, the possibility is ruled out for any of the ECs to be a variable. Some of the ECs can, however, be admitted as pronominals. The subject ECs in (52a) and (52c), in particular, can be coindexed with their respective AGR. Since the AGR is rich enough, the condition of recoverability can be satisfied, and these sentences are well-formed. The object ECs in (52b) and (52d), however, cannot be admitted as pronominals, because they would each entail a contradiction of the GCR and the binding theory governing pronominal disjoint

reference, as before. Finally, the EC in (52e) is admissible as a pronominal as in English, because it can be controlled by the matrix subject in accordance with the GCR.

Turning now to Chinese-type languages, note that because these are by assumption empty-topic languages, there are two potential ways to admit each of the ECs. Consider the possibility of each EC being a pronominal first. The subject EC in (52a) cannot be a pronominal because there is no nominal element around to satisfy the GCR. The object EC in (52b) and that in (52d) cannot be a pronominal either, because of a potential contradiction between the GCR and the binding theory. The subject EC in (52c), however, can be admitted. This is because there is no AGR in the embedded finite clause in Chinese, and in this case the closest nominal element is the matrix subject. The EC can be controlled by *John*, as is the EC in (52e). Thus, the subject EC of a finite clause in Chinese is treated on a par with the subject EC of a non-finite clause. In short, the ECs in (52c) and (52e) can be admitted as empty pronominals, but not those in (52a), (52b) and (52d).

Consider finally the possibility of each EC being a variable bound by an empty operator in Chinese. No known principle prevents the ECs in (52a), (52b), and (52d) as variables. Therefore, all of these sentences are well-formed, with each EC interpreted as referring to a discourse topic. There is also no reason why the EC in (52c) cannot be also a variable (in addition to being a pronominal controlled by *John*). Therefore, the sentence also allows such an interpretation, with the EC referring to some discourse topic. (The ambiguity of (52c), in other words, arises from the dual status of the embedded subject EC.) Finally, although the EC in (52e) can be interpreted as being bound by *John*, it cannot be interpreted otherwise. This means that it cannot be admitted as a variable referring to a discourse topic. This situation is not unexpected, however: Given that the embedded sentence is non-finite, the EC is ungoverned. The ECP thus independently rules out the possibility of it being a variable in this position.

Summarizing, the distribution of empty pronominals is accounted for in the following way. Such empty elements are not admitted in object position because of the interaction of the GCR and the binding theory. In subject position a pronominal EC is admitted when there is a rich enough AGR, in which case its reference is determined by the AGR, and when there is no AGR at all, in which case its reference is determined by a controller NP. The one situation where such an EC is not allowed is when there is a meager AGR (as in English). The mere presence of AGR makes it the necessary antecedent of the EC in accordance of the GCR. But its degenerate nature prevents it from fulfilling the condition of recoverability.

## REFERENCES

- Borer, H. (1983) *Parametric Syntax*, Foris Publications, Dordrecht.
- Chomsky, N. (1980) 'On Binding,' *Linguistic Inquiry* 11,1-46.
- \_\_\_\_\_. (1981) *Lectures on Government and Binding*, Foris Publications, Dordrecht.
- \_\_\_\_\_. (1982) *Some Concepts and Consequences of the Theory of Government and Binding*, MIT Press, Cambridge, Massachusetts.
- Evans, G. (1980) 'Pronouns,' *Linguistic Inquiry* 11,337-362.
- Gundel, J. (1980) 'Zero Anaphora in Russian: A Case of Topic Prominence,' *Proceedings of the Parasession on Pronouns and Anaphora*, Chicago Linguistic Society, 139-146.
- Huang, C. T. J. (1982) *Logical Relations in Chinese and the Theory of Grammar*, doctoral dissertation, MIT.
- \_\_\_\_\_. (1984) 'On the Distribution and Reference of Empty Pronouns,' *Linguistic Inquiry* 15.4.
- Jaeggli, O. (1982) *Topics in Romance Syntax*, Foris Publications, Dordrecht.
- Lasnik, H. (1976) 'Remarks on Coreference,' *Linguistic Analysis* 2,1-22.
- Li, C. and S. Thompson (1976) 'Subject and Topic: A New Typology,' in C. Li, ed., *Subject and Topic*, Academic Press, New York.
- May, R. (1977) *The Grammar of Quantification*, doctoral dissertation, MIT.
- McCloskey, J. and K. Hale (1983) 'The Syntax of Inflection in Modern Irish,' *Proceedings of the 13th Conference of the North Eastern Linguistic Society*, 173-190.
- Perlmutter, D. (1971) *Deep and Surface Constraints in Syntax*, Holt, Rinehart, and Winston, New York.
- Picalo, C. (1982) 'The INFL Node and the Pro Drop Parameter,' mimeographed, MIT and CUNY Graduate Center.
- Rizzi, L. (1982) *Issues in Italian Syntax*, Foris Publications, Dordrecht.
- Rosenbaum, P. S. (1967) *The Grammar of English Predicate Complement Constructions*, MIT Press, Cambridge, Massachusetts.
- Ross, J. R. (1982) 'Pronoun Deleting Processes in German,' paper presented at the annual meeting of the Linguistic Society of America, San Diego, California.
- Taraldsen, T. (1978) 'On the NIC, Vacuous Application, and the That-trace Filter,' Indiana University Linguistics Club.
- Tsao, F. (1977) *A Functional Study of Topic in Chinese: The First Step toward Discourse Analysis*, doctoral dissertation, USC.
- Xu, L. J. (1984) 'Empty Categories in Chinese,' GLOW newsletter, February 1984.
- Yang, D. W. (1983) 'The Extended Binding Theory for Anaphors,' paper presented at GLOW, University of York.
- Zubizarreta, M.-L. 'Theoretical Implications of Subject Extraction in Portuguese,' *The Linguistic Review* 2,79-96.

Department of Chinese Literature and Linguistics, *and*  
Department of Foreign Languages  
National Tsing Hua University  
Hsinchu 300  
Taiwan, R.O.C.